Recent Probabilistic and Computational Advances in Dose-Response Assessment to Better Support Chemical Risk Assessment

March 3, 2020
SRA/DRSG Webinar Series

Weihsueh A. Chiu, PhD
Texas A&M University

Acknowledgments

This work was made possible, in part, by NIH grant P42 ES027704 and KS and Associates LLC (via NIH STTR 1R41TR002567-01A1). Its contents are solely the responsibility of the grantee and do not necessarily represent the official views of the NIH. Further, the NIH does not endorse the purchase of any commercial products or services mentioned in the publication. The views in this article are those of the authors and do not necessarily reflect the views or policies of the NIH. The authors declare they have no actual or potential competing financial interests.

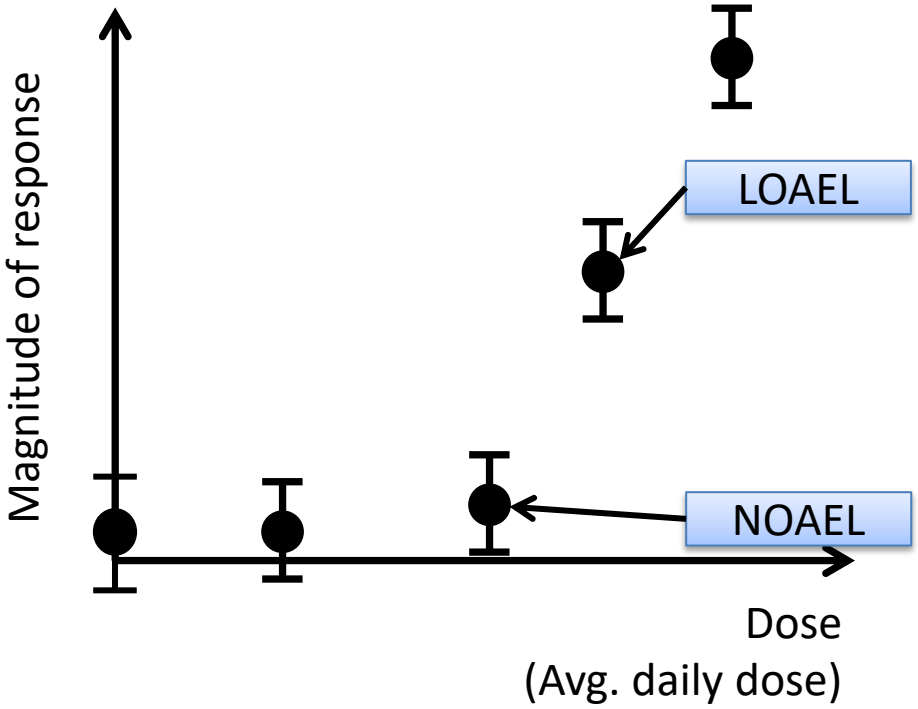
Daniel A. Axelrad, U.S. EPA
Lesa Aylward, Summit Toxicology
David Bussard, U.S. EPA
Chimeddulam Dalaijamts (postdoc), TAMU
Chris Dockins, U.S. EPA
Peter Fantke, Tech. U. Denmark
Dale Hattis, Clark University
Matthias Herzler, BfR
Suji Jang (grad student), TAMU
Olivier Jolliet, U. Michigan
Richard Judson, U.S. EPA
Kan Shao, Indiana University
Andrew J. Shapiro, U.S. EPA
Wout Slob, RIVM
Woody Setzer, U.S. EPA (retired)
Lauren Zeise, California EPA

Outline

- Motivation and background
- Overview of the WHO/IPCS framework
- Recent advances in apply the WHO/IPCS framework
- Remaining challenges and summary

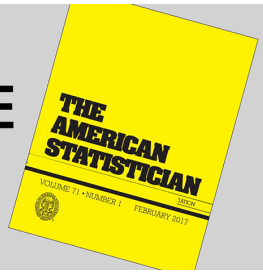
Limitations of the NOAEL

- Limited to doses in the study
- Minimum detectable response increases as sample size decreases
- Role of expert judgment often unclear
- Not comparable across studies
- Does not characterize uncertainty or variability



Concept of NOAEL based on statistical significance is ***fundamentally flawed***

SPECIAL ISSUE ON *P*-VALUE



Moving to a World Beyond "$p < 0.05$"

Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein, Allen L. Schirm & Nicole A. Lazar (2019) Moving to a World Beyond "$p<0.05$", The American Statistician, 73:sup1, 1-19, DOI: 10.1080/00031305.2019.1583913

43 additional papers…

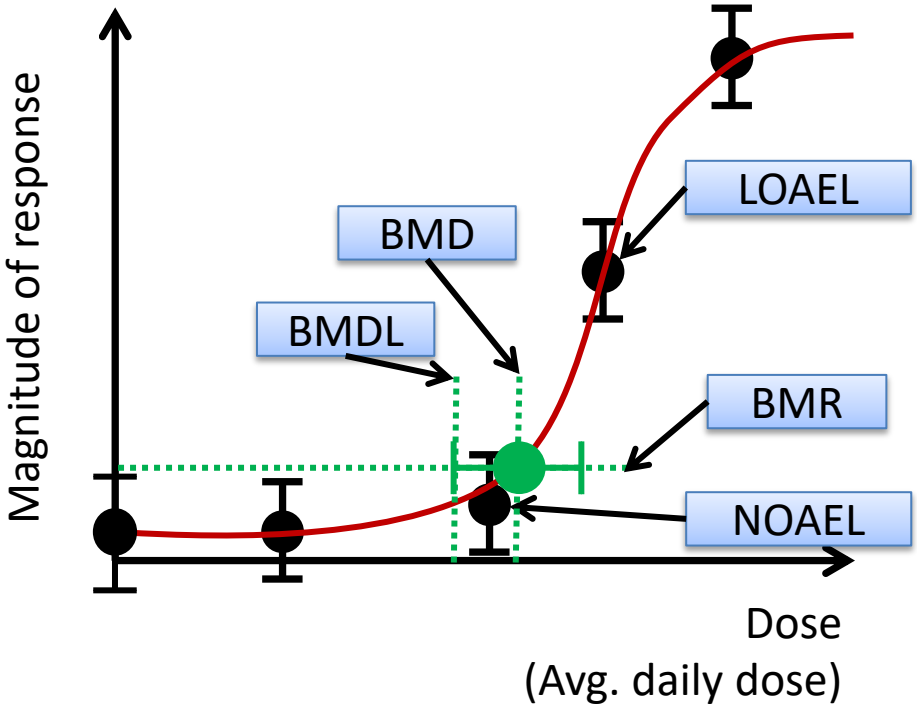
NOAEL, as practiced, violates ALL of these "don't"s

There's not much we can say here about the perils of *p*-values and significance testing that hasn't been said already for decades (Ziliak and McCloskey 2008; Hubbard 2016). If you're just arriving to the debate, here's a sampling of what not to do:

- Don't base your conclusions solely on whether an association or effect was found to be "statistically significant" (i.e., the *p*-value passed some arbitrary threshold such as $p < 0.05$).
- Don't believe that an association or effect exists just because it was statistically significant.
- Don't believe that an association or effect is absent just because it was not statistically significant.
- Don't believe that your *p*-value gives the probability that chance alone produced the observed association or effect or the probability that your test hypothesis is true.
- Don't conclude anything about scientific or practical importance based on statistical significance (or lack thereof).

Benefits of Benchmark Dose*

- Fitting a curve so not limited to doses in the study
- Magnitude of response fixed at a "benchmark response" (BMR) level
  - Clarifies role of expert judgment
  - Avoids misinterpretation as "threshold"
  - Aids comparability across studies
- **Accepts uncertainty,** and characterizes it in a confidence interval
  - Sample size
  - Experimental variability



*Crump (1984)

Benefits are a consequence of benchmark dose having a more precise definition than the NOAEL

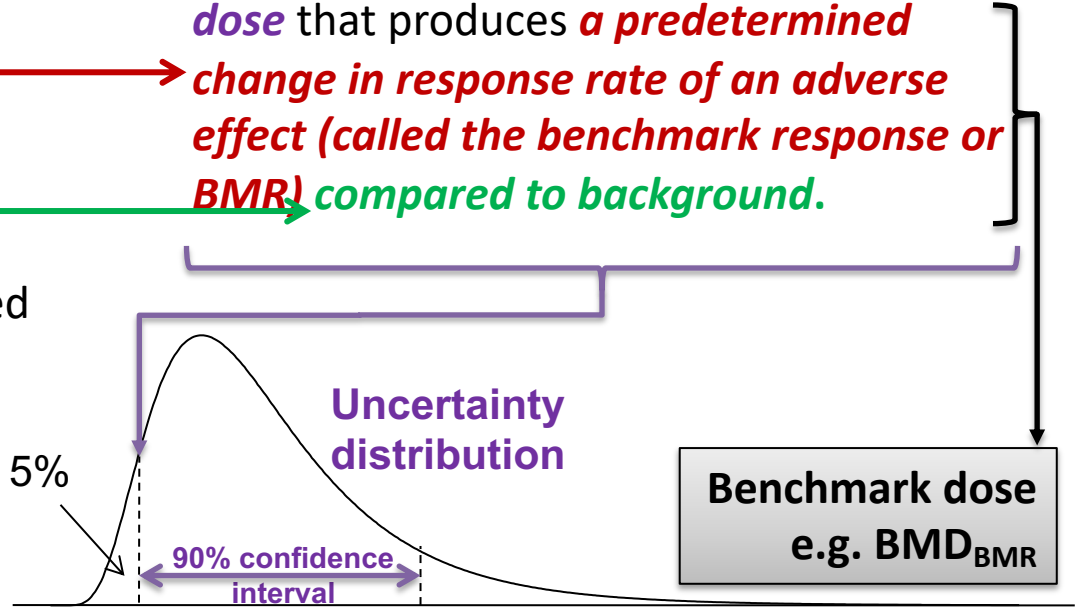
NOAEL:

Greatest concentration or amount of a substance, found by experiment or observation, that causes ***no adverse alteration …of the target organism distinguishable from those observed in normal (control) organisms*** of the same species and strain under the same defined conditions of exposure.

NOAEL should be viewed as an "approximation" of the BMD!

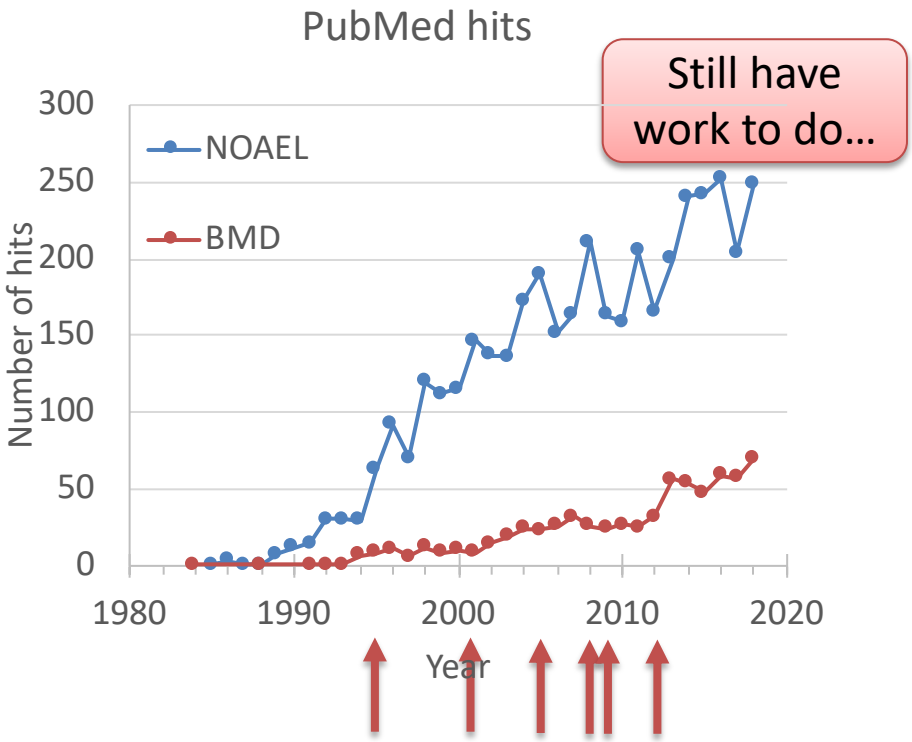
BMDL:

A statistical lower confidence limit on the dose that produces ***a predetermined change in response rate of an adverse effect (called the benchmark response or BMR) compared to background.***



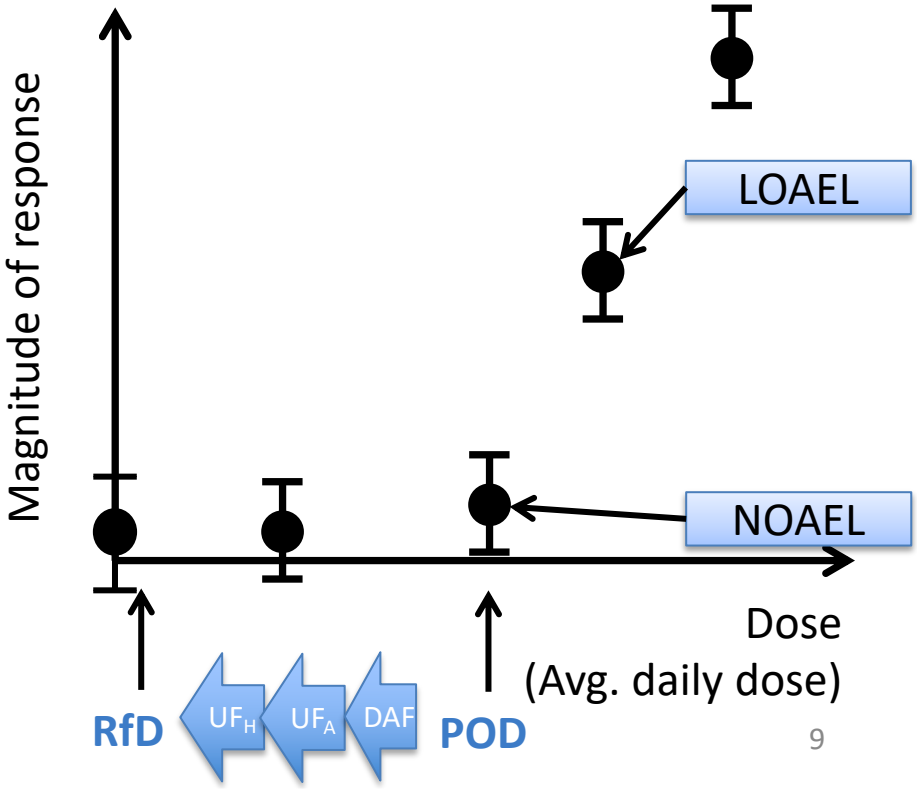
Generally Agreed that Benchmark Dose Is More Scientifically Valid than NOAEL

- U.S. EPA Risk Assessment Forum Report on The Use of the Benchmark Dose Approach in Health Risk Assessment (1995)
- National Research Council Standard Operating Procedures for Developing Acute Exposure Guideline Levels (2001)
- Joint FAO/WHO Meeting on Pesticide Residues (2005)
- European Chemicals Agency Guidance (2008)
- European Food Safety Authority Scientific Opinion (2009)
- World Health Organization Principles for modeling dose-response (2009)
- U.S. EPA Benchmark Dose Technical Guidance (2012)



Similar Issues with the deterministic RfD!

- **POD (Point of Departure):** If NOAEL, then ***magnitude of effect*** is unspecified and ***uncertainty*** is unquantified
- **DAF (Dosimetric Adjustment Factor):** Accounts for "average" interspecies differences (e.g., allometric scaling), uncertainty unquantified
- **UF_A (Interspecies Uncertainty Factor):** Assumed to be ***conservative***, but unclear by how much
- **UF_H (Intraspecies Uncertainty Factor):** Factor accounting for ***variability*** assumed to be ***conservative***, but unclear by how much and unspecified as to ***population fraction*** covered
- **RfD (Reference Dose):**
  - Assumed to be ***conservative***, but unclear by how much and unspecified as to ***population fraction*** covered
  - More ***conservative*** with more factors
  - Different assessments may differ greatly in level of ***conservatism***
  - Just providing a ***conservative*** bound not optimal for many types of decisions



Issues Recognized by the National Academies

- ***Science and Judgment*** report (NRC, 1994) recommended presenting quantitative representation of uncertainty.

- ***Science and Decisions*** report (NRC, 2009) recommended incorporating
  - Mode of action, vulnerable populations, background exposures
  - Unified approach to both cancer and non-cancer endpoints
  - Probabilistic methods for assessing uncertainty.

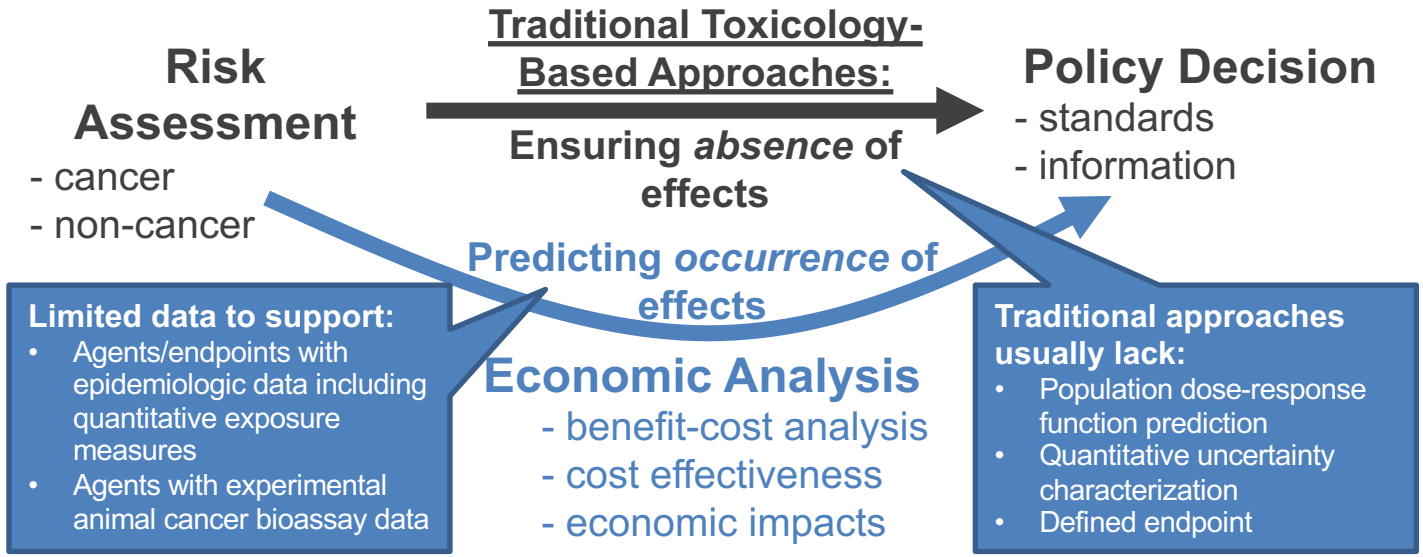
- ***Review of the IRIS Program*** report (NRC, 2014) recommended systematic use of uncertainty analysis and expanded use of Bayesian methods.

Existing RfD is ill-suited for many decision contexts

- Most chemical “risk assessments” are actually “safety assessments”
  - Risk: likelihood or probability of adverse effect(s)
  - Safety: providing “assurance” of an “absence” of risk

- Socio-economic benefits analyses require
  - Economically meaningful health effect(s)
  - Expected change in number of cases under different policy options

An Economist's View of Toxicology and Risk Assessment

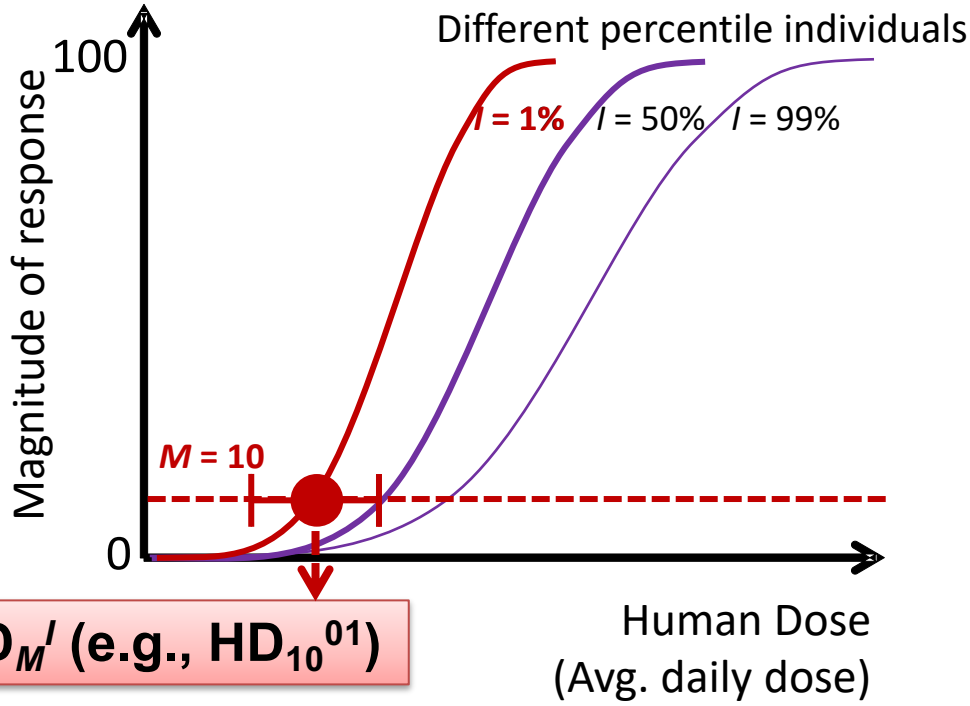


Outline

- Motivation and background
- **Overview of the WHO/IPCS framework**
- Recent advances in apply the WHO/IPCS framework
- Remaining challenges and summary

WHO approach to go “Beyond the RfD” to the “Target Human Dose”

- **Target Human Dose (HD_M^I):** human dose that at which a ***fraction I*** of the population shows an effect of ***magnitude (or severity) M*** or greater (for the critical effect considered).
- Specifies the “target” ***magnitude of effect M*** (analogous to BMR for the benchmark dose)
- Specifies “target” ***fraction of the variable population I (incidence)***
- Can be estimated probabilistically to derive a ***confidence interval*** that characterizes ***uncertainty***.



Key Concept: HD_M^I (e.g., HD_{10}^{01})

WHO (2014): Guidance on Evaluating and Expressing Uncertainty in Hazard Assessment. Harmonization Project Document 11.
Chiu WA & Slob W (2015): A Unified Probabilistic Framework for Dose-Response Assessment of Human Health Effects. EHP, DOI: 10.1289/ehp.1409385

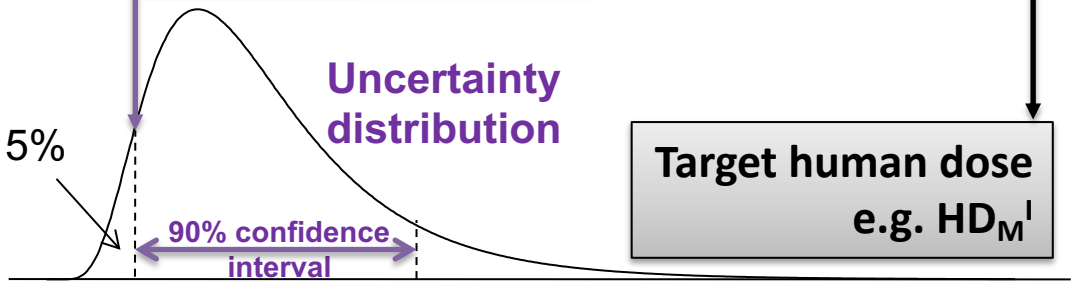
Target Human Dose (HD_M^I) and Probabilistic RfD have more precise definitions than the RfD

RfD:

An estimate (with uncertainty spanning perhaps an order of magnitude) of a daily oral exposure to the ***human population (including sensitive subgroups)*** that is ***likely*** to be ***without an appreciable risk*** of ***deleterious effects*** during a lifetime.

Probabilistic RfD:

A statistical lower confidence limit on the human dose that at which ***a fraction I of the population*** shows ***an effect of magnitude (or severity) M or greater (for the critical effect considered)***.



RfD should be viewed as an "approximation" of the HD_M^I!

Exactly analogous to transition from NOAEL to BMDL (*"déjà vu all over again…"*)

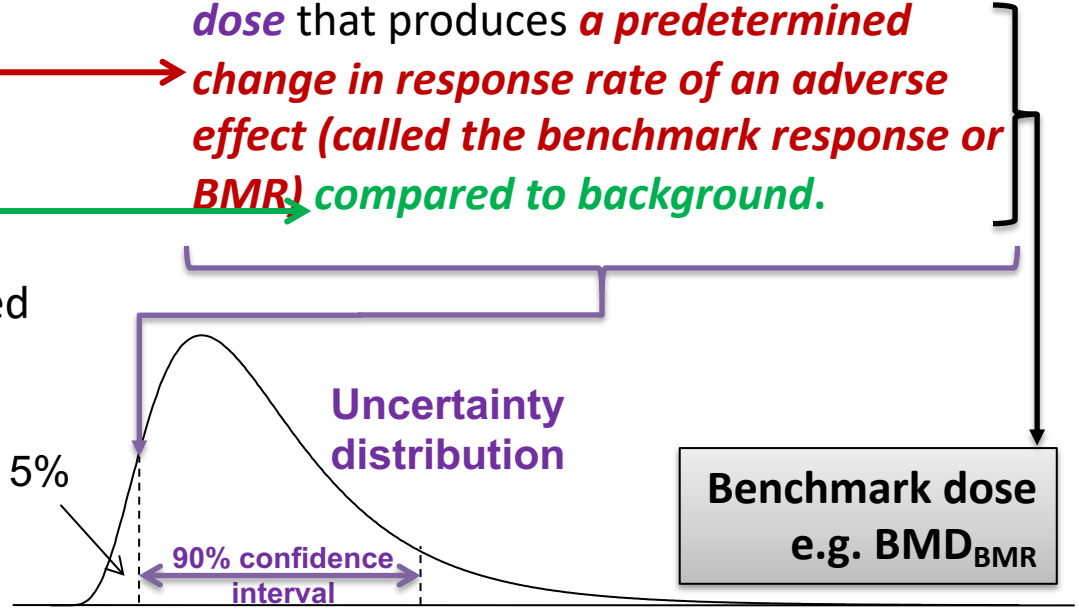
NOAEL:

Greatest concentration or amount of a substance, found by experiment or observation, that causes ***no adverse alteration …of the target organism distinguishable from those observed in normal (control) organisms*** of the same species and strain under the same defined conditions of exposure.

BMDL:

A statistical lower confidence limit on the dose that produces ***a predetermined change in response rate of an adverse effect (called the benchmark response or BMR) compared to background.***

NOAEL should be viewed as an "approximation" of the BMD!

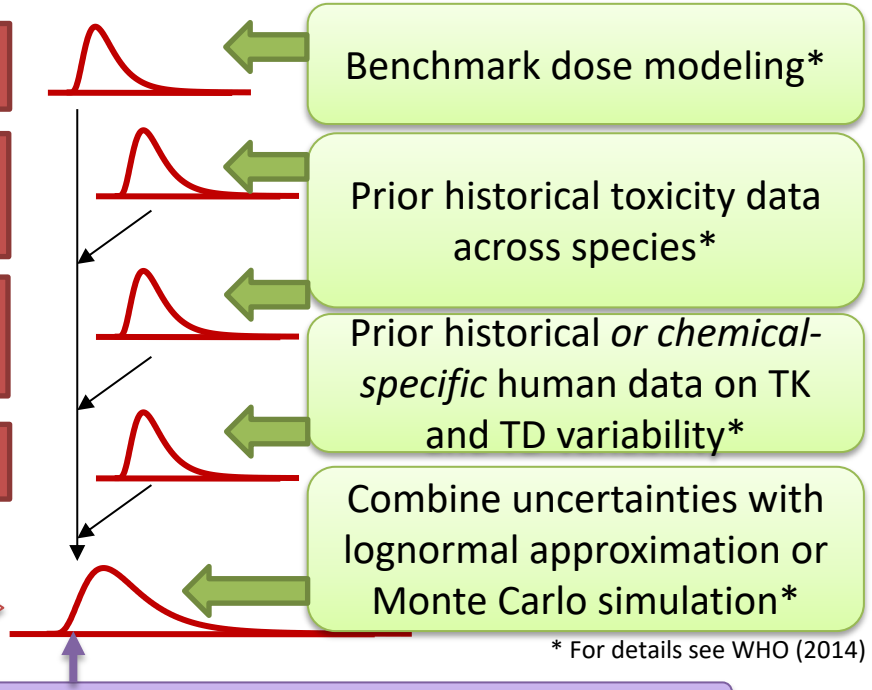


Target Human Dose Derived Similarly to a RfD, but Addressing Its Limitations

$$RfD = \frac{NOAEL \times DAF}{UF_A \times UF_H}$$

$$HD_M^I = \frac{BMD_M \times DAF}{UF_A \times UF_{H,I}}$$

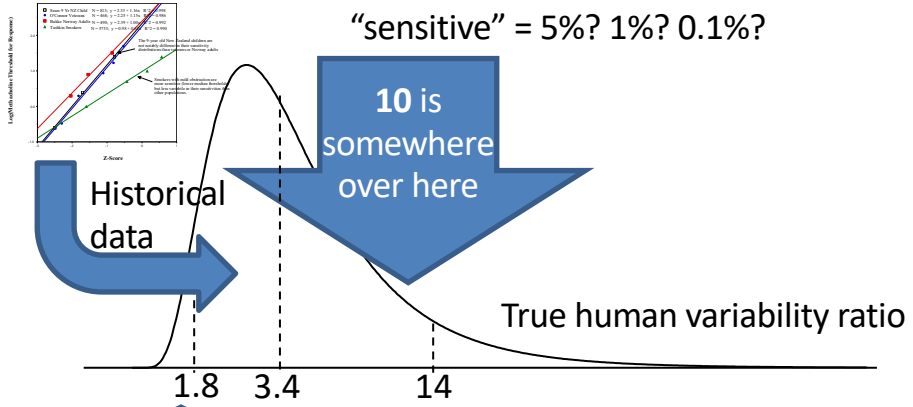
- Benchmark Dose for Magnitude of effect M
- Dosimetric Adjustment Factor for Body Size differences
- Interspecies Factor for remaining TK and TD differences
- Human Variability Factor for Incidence I



* For details see WHO (2014)

Lower Confidence Bound can be defined as a "Probabilistic RfD"

Revisiting "generic" uncertainty factors



(By definition of "sensitive," has to be >1)

Historical data on human variability:

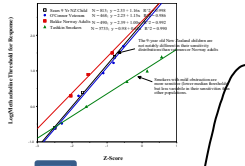
For "sensitive" = 5%

- Median ~ 3.4
- 95th %ile ~ 14
- 10 corresponds to ~90%ile

- What does a human variability factor of 10 mean?
  - The range from "least sensitive" to "most sensitive" humans is 10-fold.
  - "Sensitive" humans are 10x more sensitive than "typical" humans.
  - For most chemicals and endpoints, "sensitive" humans are no more than 10x more sensitive than "typical" humans.
- Thus, the TK and TD factors of 3 are intended to cover "most" chemicals for "sensitive" humans.

Reanalysis of Hattis and Lynch (2007) data by WHO/IPCS (2014).

Revisiting "generic" uncertainty factors



"most" = 90%? 95%? 99%?
"sensitive" = 5%? 1%? 0.1%?

10 is somewhere over here

Historical data

True human variability ratio

2.2 5.7 42

1.1 is somewhere over here

(By definition of "sensitive," has to be >1)

Historical data on human variability:

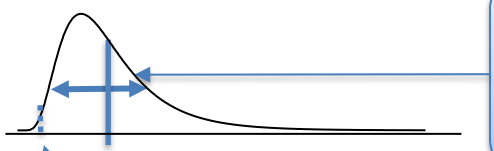
For "sensitive" = 1%

- Median ~ 5.7
- 95th %ile ~ 42
- 10 corresponds to ~70%ile

- What does a human variability factor of 10 mean?
  - The range from "least sensitive" to "most sensitive" humans is 10-fold.
  - "Sensitive" humans are 10x more sensitive than "typical" humans.
  - For most chemicals and endpoints, "sensitive" humans are no more than 10x more sensitive than "typical" humans.
- Thus, the TK and TD factors of 3 are intended to cover "most" chemicals for "sensitive" humans.

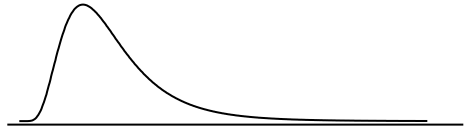
Reanalysis of Hattis and Lynch (2007) data by WHO/IPCS (2014).

Human variability: Disaggregating Uncertainty and Variability



Human equipotent dose

Variability distribution is approximately lognormal: **Interindividual variability, GSD_H**



Estimates of **Log(GSD_H)** for different chemicals

*l*th percentile defines "sensitive" individual being protected (e.g., 1st percentile)

$$UF_{Intra\text{-}l} = GSD_H^{|z_l|}$$

Uncertainty distribution is approximately lognormal: **Chemical-specific uncertainty, GSD_U**

Uncertainty distribution for $UF_{Intra\text{-}l}$

(depends on choice of *l*!)

Combining Uncertainties

APROBA & APROBAweb

- Uses lognormal distribution for each *factor*
- BMD approximated by lognormal with lower 5%ile=BMDL and either
  - median=BMD or
  - upper 95%ile=BMDU
- DAF & UF_A are already assumed to be lognormal
- $UF_{H,I} = GSD_H^{|z|}$ approximated by lognormal (equivalent to $\log(GSD_H)$ approximated by normal instead of being lognormal)
- Result is lognormal HD_M^I.

$$HD_M^I = \frac{BMD_M \times DAF}{UF_A \times UF_{H,I}}$$

Bayesian BMD

incorporates Chiu & Slob 2015 R code

- BMD distribution uses samples from Bayesian posterior
- Incorporates Chiu & Slob (2015) R code for MC simulation for remaining factors
- DAF & UF_A are sampled from their assumed lognormal distributions
- $UF_{H,I} = GSD_H^{|z|}$ samples drawn based on assumed lognormal distribution of $\log(GSD_H)$
- Result is set of samples for HD_M^I.

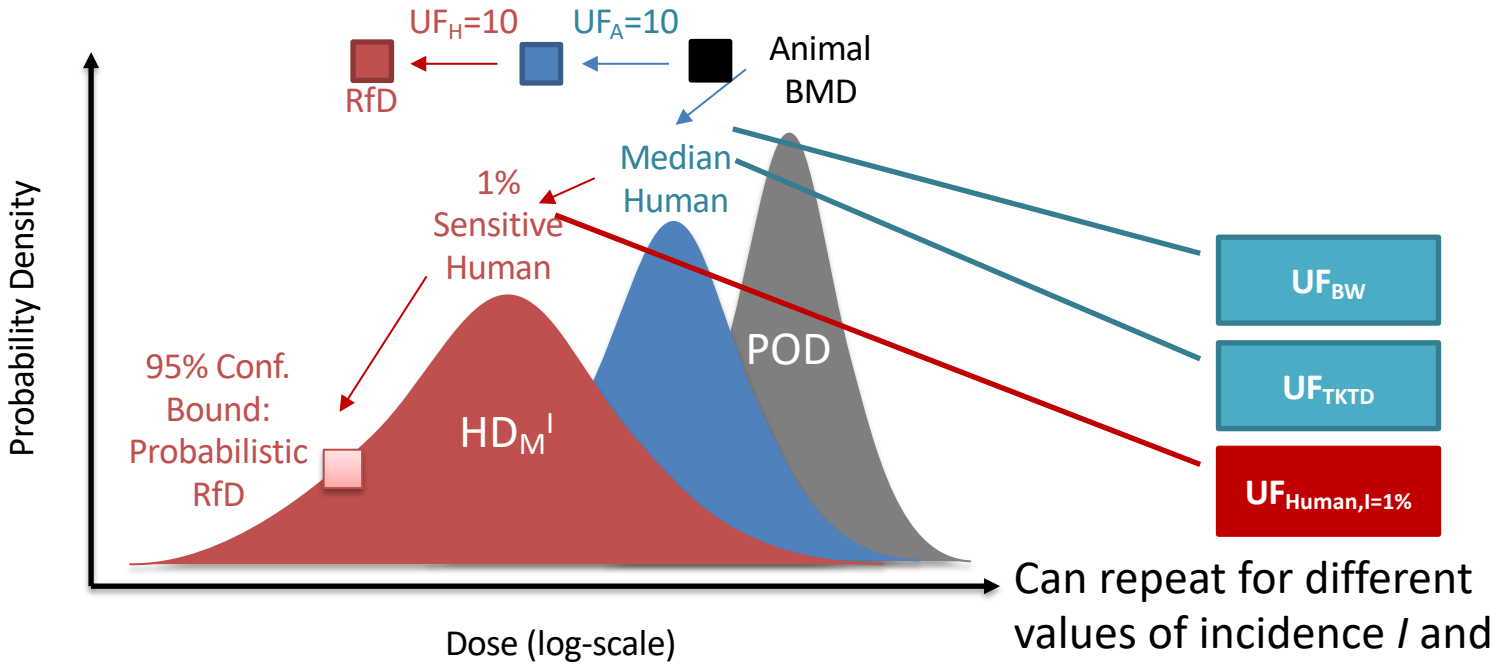
APROBAweb Interactive Web Application (based on Rshiny)
https://wchiu.shinyapps.io/APROBAweb/

Web-based Bayesian Benchmark Dose (BBMD) System
https://benchmarkdose.org/ (Shao & Shapiro 2018)

Illustration of HD_M^I Calculation

Probabilistic RfD:

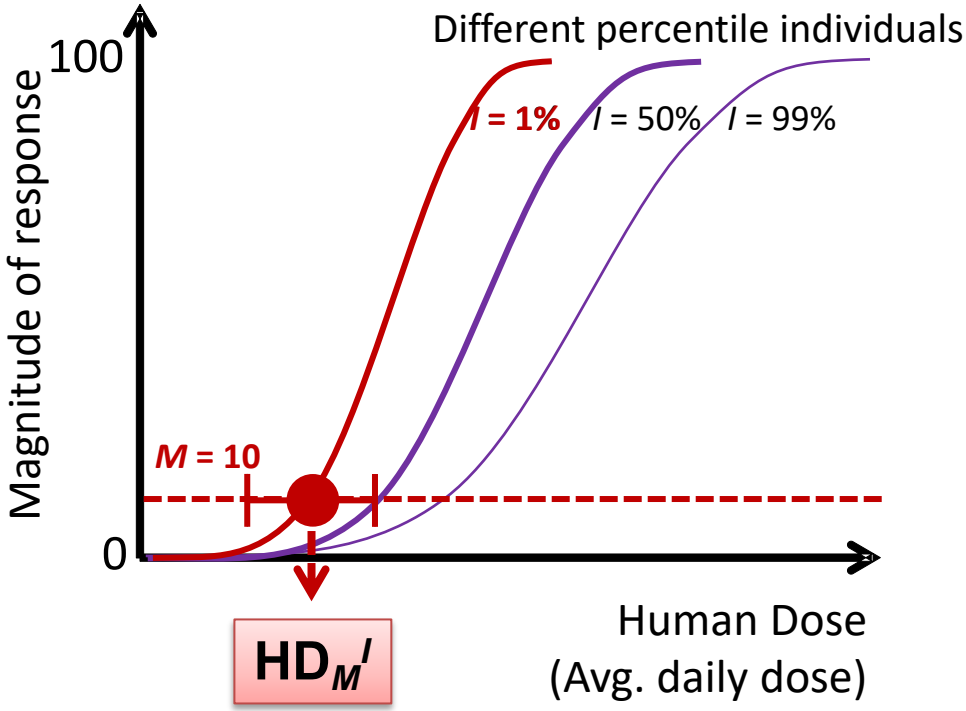
A statistical lower confidence limit on the human dose that at which ***a fraction I of the population*** shows ***an effect of magnitude (or severity) M or greater (for the critical effect considered)***.



Can repeat for different values of incidence *I* and magnitude *M*

Applications of Probabilistic Dose-Response

- Can **replace** RfD or RfC.
- Addresses **population variability** through ***I***.
- Provides a **dose-response function** through ***M***.
  - Effects of incremental changes in dose
  - Cumulative risk of multiple chemicals
- Characterizes **uncertainty**
- Can incorporate **chemical-specific data** on TK and TD variability



Outline

- Motivation and background
- Overview of the WHO/IPCS framework
- **Recent advances in apply the WHO/IPCS framework**
- Remaining challenges and summary

Demonstration of Feasibility

- **Identification and curation of chemical RfDs**
  - Use only RfDs from federal health agencies
  - Wide range of effect types, PODs, composite UFs
- **Automated Workflow for Probabilistic dose-response assessment**
  - Convert to endpoint-specific RfDs
  - Assign conceptual models and uncertainty distributions
  - Combine uncertainties probabilistically

Demonstrating Feasibility

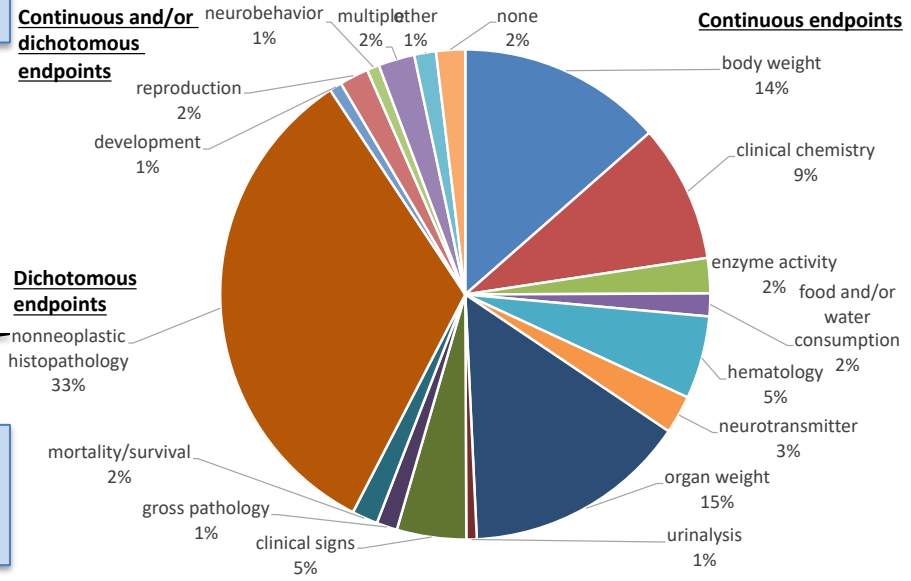
3064 peer-reviewed toxicity values and endpoints from Wignall et al. (2014)

- Non-cancer, oral RfDs
- Based on animal data
- Sufficient reporting of species, points of departure, uncertainty factors, etc.

1464 RfDs and endpoints

608 Chemicals (351 with multiple RfDs or endpoints)

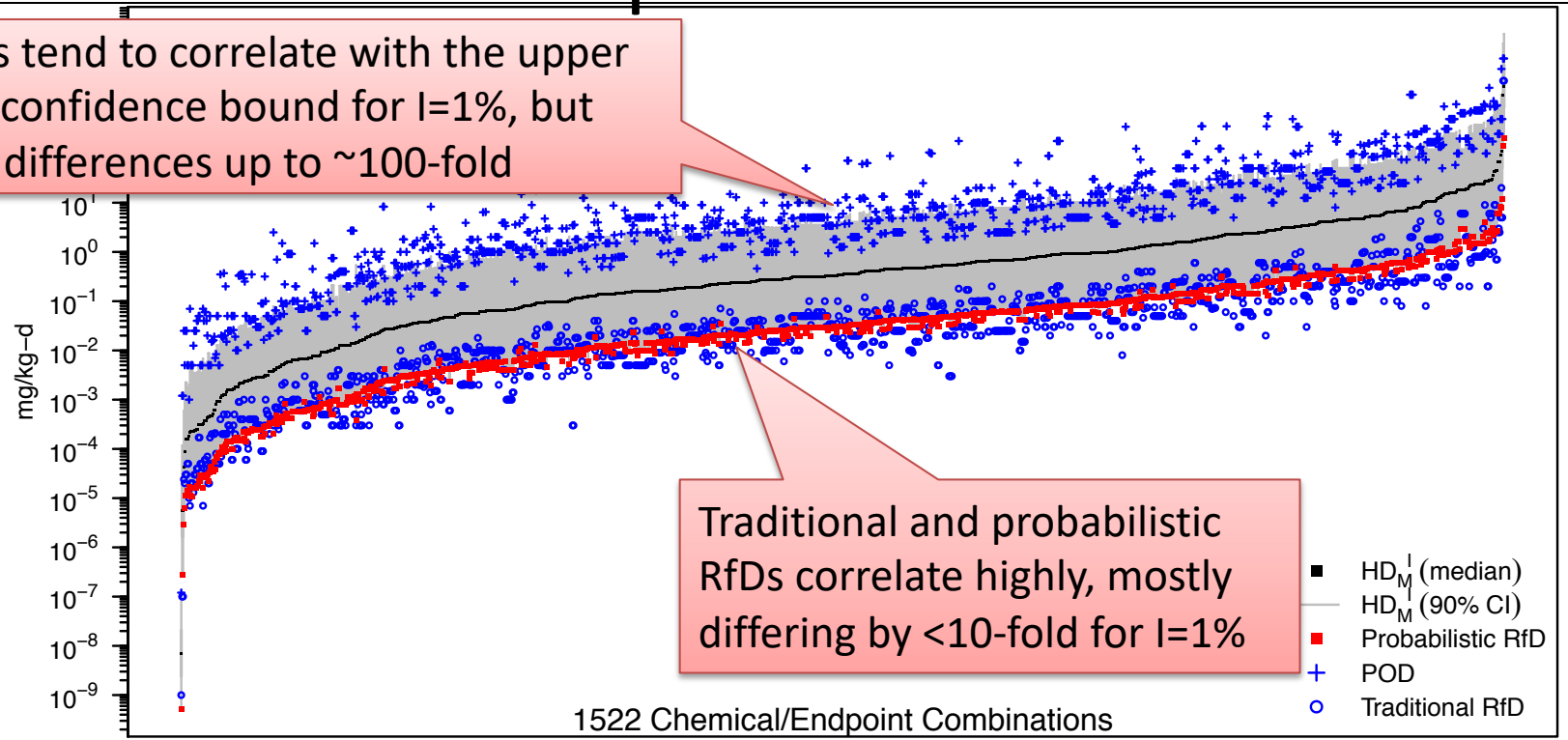
Effect Types



Re-Calculate using Probabilistic Methods

Comparing Traditional and Probabilistic Dose-Response Assessment

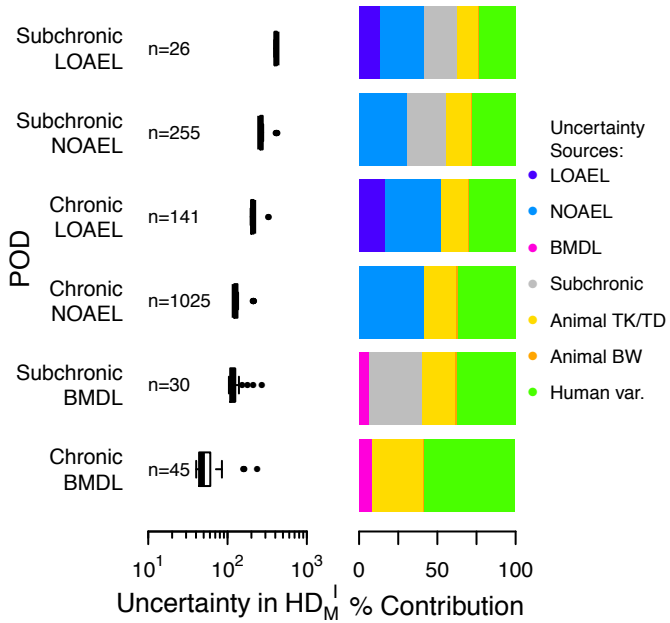
PODs tend to correlate with the upper 95% confidence bound for I=1%, but with differences up to ~100-fold



Traditional and probabilistic RfDs correlate highly, mostly differing by <10-fold for I=1%

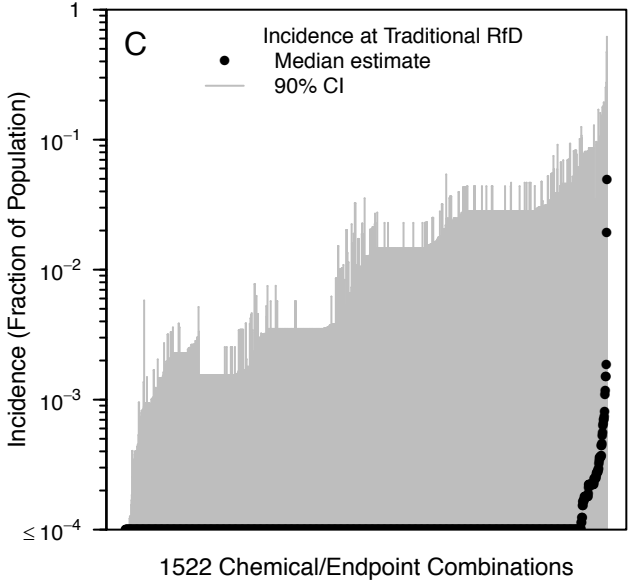
Sources of Uncertainty in HD_M^I

- Degree of uncertainty depends strongly on the POD type:
  - LOAELs from subchronic studies have greatest uncertainty (~400-fold range)
  - BMDLs from chronic studies have least uncertainty (~50-fold range).
  - Most common PODs were NOAELs from chronic studies (~100-fold range).
- Greatest contributors to uncertainty:
  - Lack of BMDL
  - Uncertainty about human variability



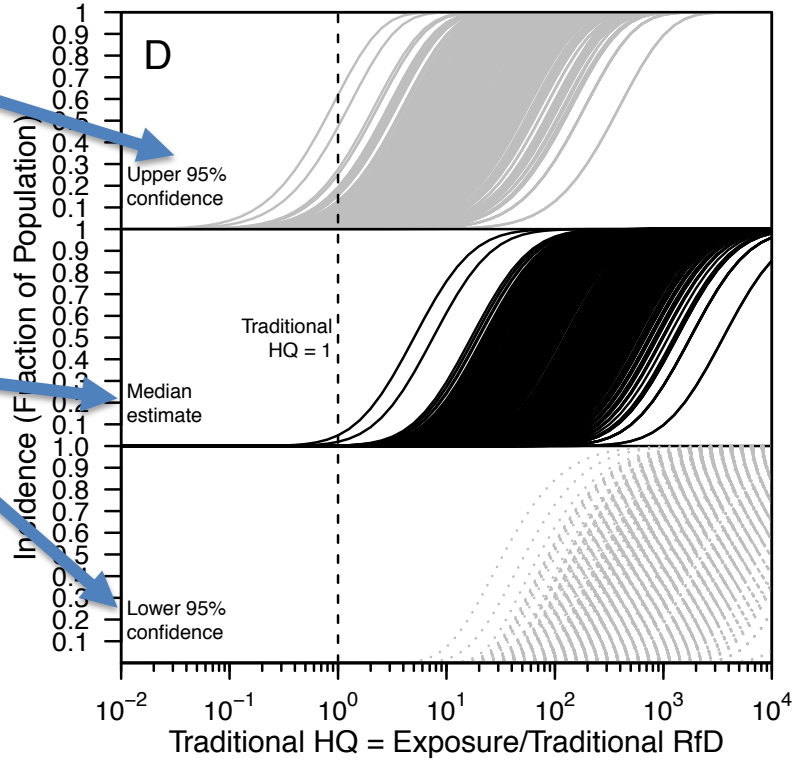
Residual Risk at Traditional RfD

- Upper bound risk typically a few percent
- Median and lower bound risks < 0.01%
- Severity of endpoints vary greatly (e.g., “mild irritation” vs. “hemorrhage”)



Generating Population Dose-Response Functions

- Upper bound risk at the traditional RfD is typically a few percent, but sometimes much more
- Median and lower bound risks mostly < 0.01%
- Can calculate expected values using Monte Carlo simulation



Chiu et al. (2018) https://doi.org/10.1289/EHP3368

Lessons Learned Across Many Chemicals and Endpoints

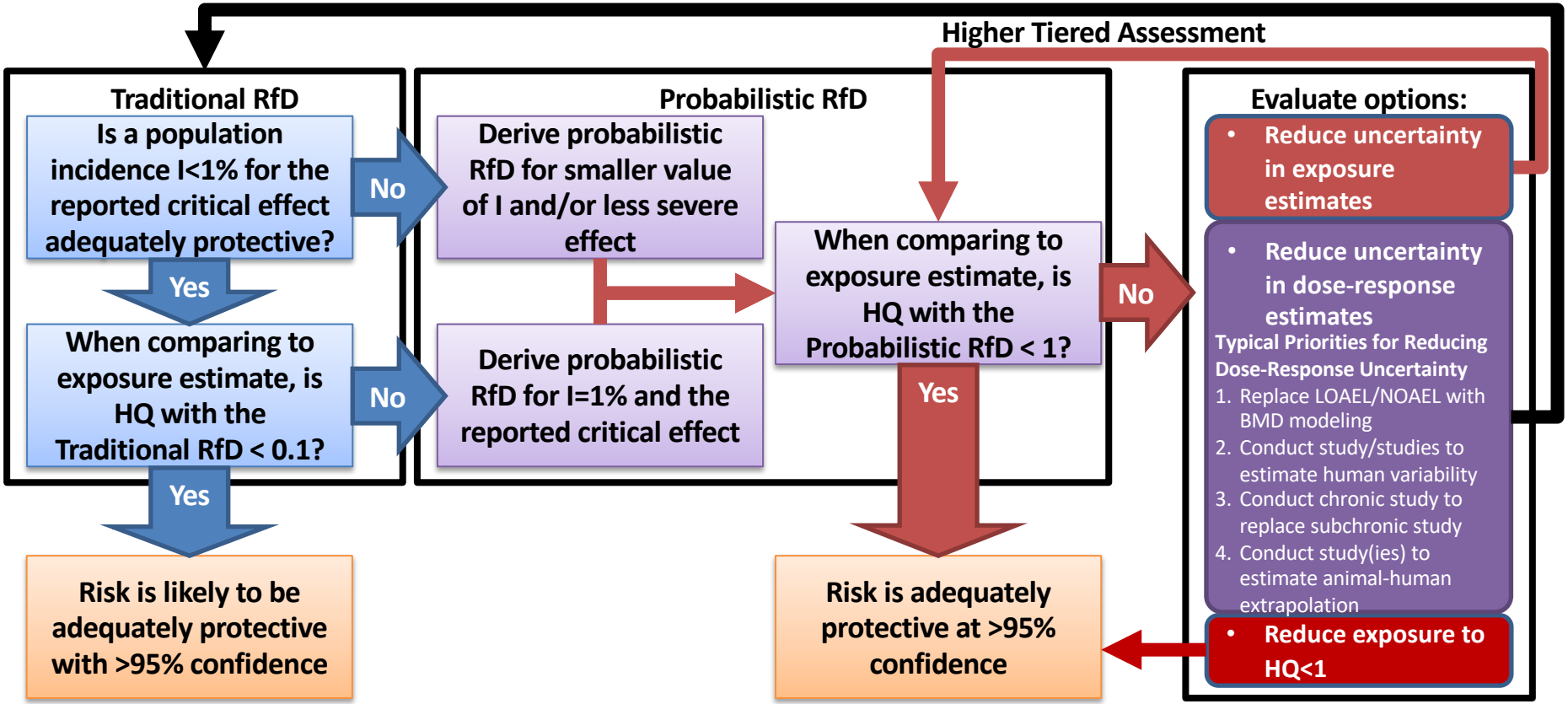
- **Broadly improving rigor, transparency, and consistency of dose-response assessment using probabilistic approach is feasible:**
  - Uncertainty factor distributions derived from historical evidence, not factors of 1, 3, or 10.
  - Combining uncertainty distributions probabilistically avoids "compounding conservatism."
  - The resulting HD_M^I is clear as to the degree of health protection (target incidence and magnitude of effect) and conservatism (% confidence)

- **Exposure at the current RfD**
  - Frequently implies upper 95% confidence bound incidence of a few percent
  - Whether such risks are "acceptable" may vary by context (including endpoint severity).

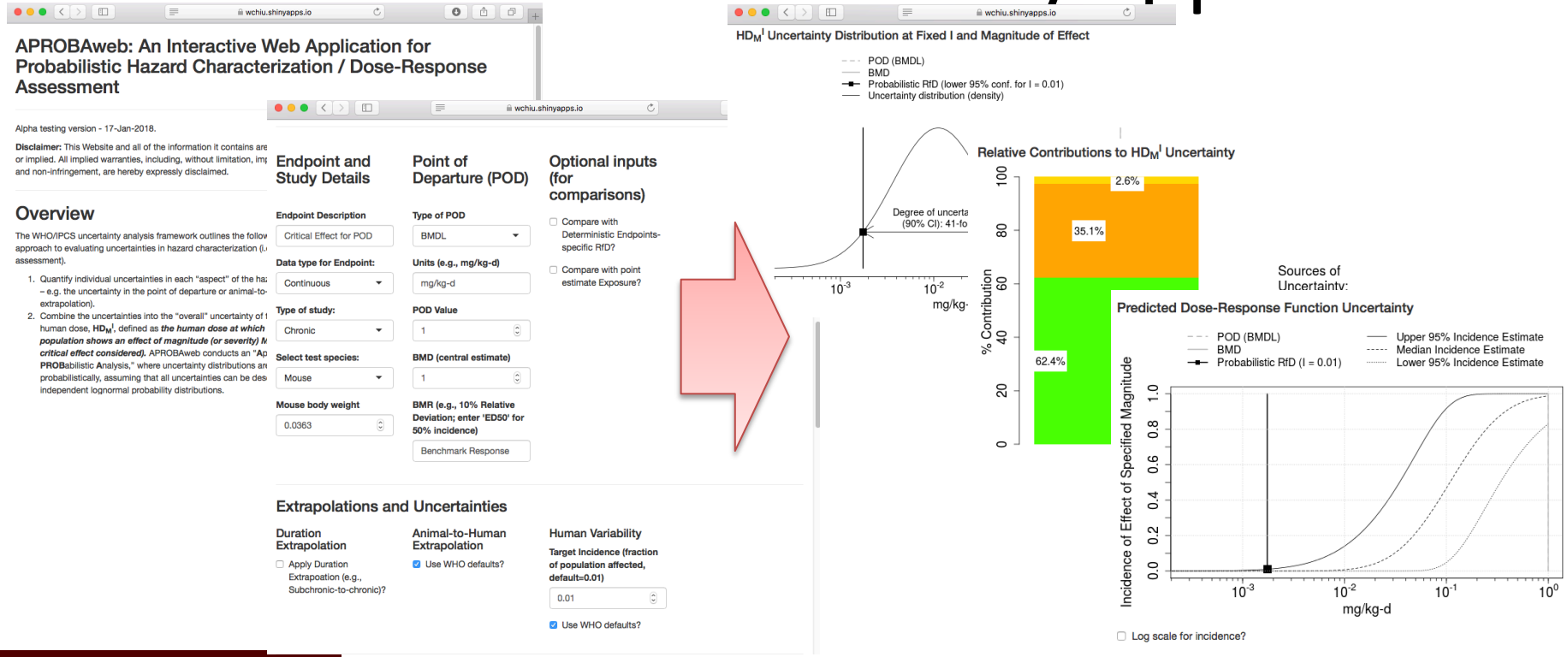
- **Broader use of BMD modeling can substantially reduce uncertainty.**

- **Research into the extent of human variability may warrant more emphasis than inter-species differences.**

Example Risk Assessment/Risk Management Workflow Incorporating Traditional RfD, Probabilistic RfD, and Tiered Uncertainty Reduction



Online Tools to Facilitate Computation: APROBAweb Rshiny app

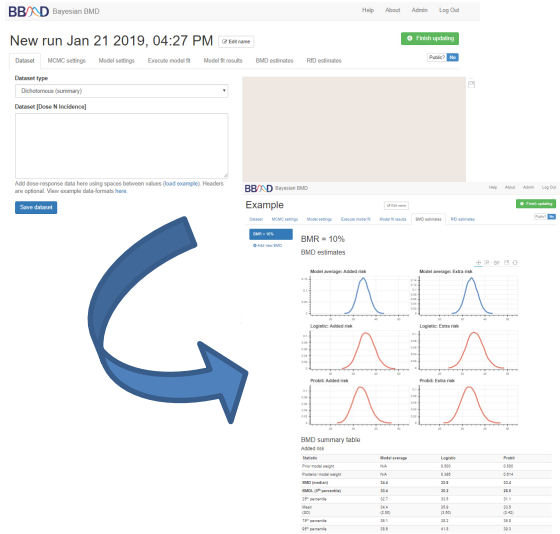


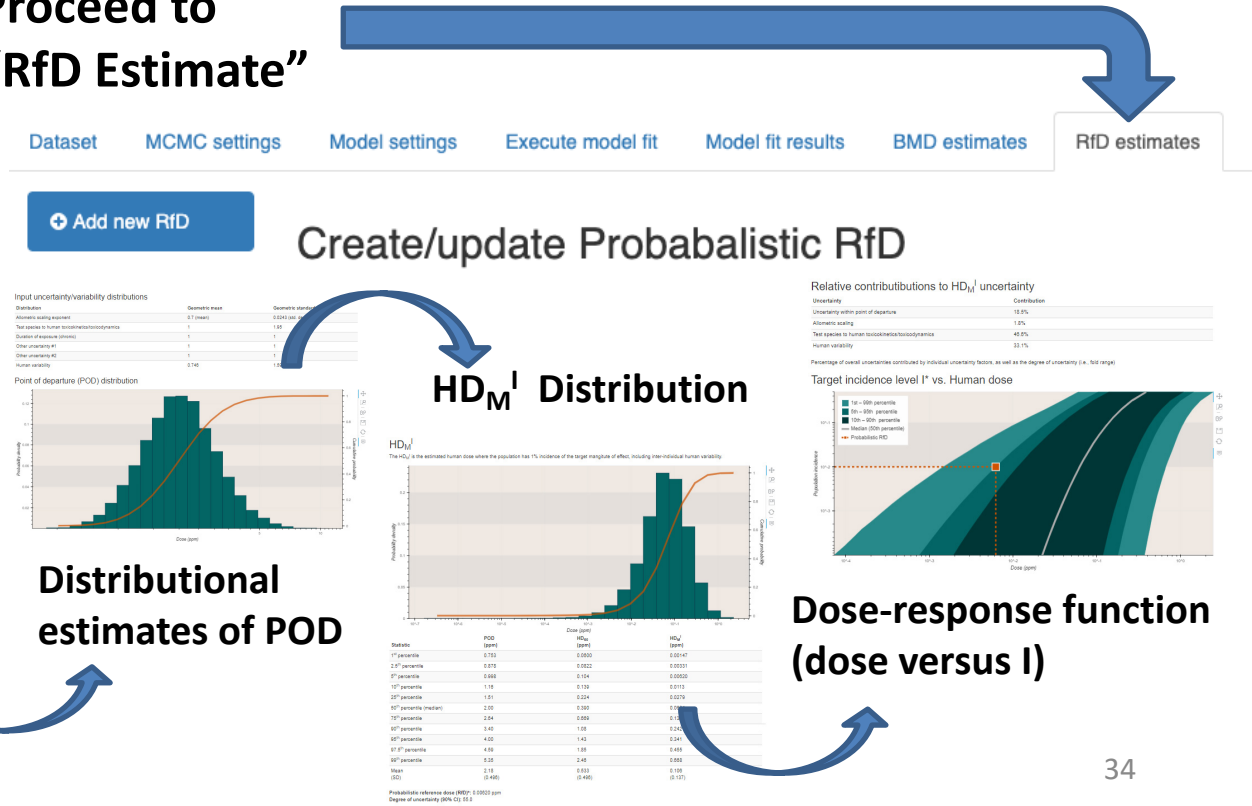
https://wchiu.shinyapps.io/APROBAweb/

Online Tools to Facilitate Computation: Bayesian Benchmark Dose Web Portal

BBMD analysis of datasets

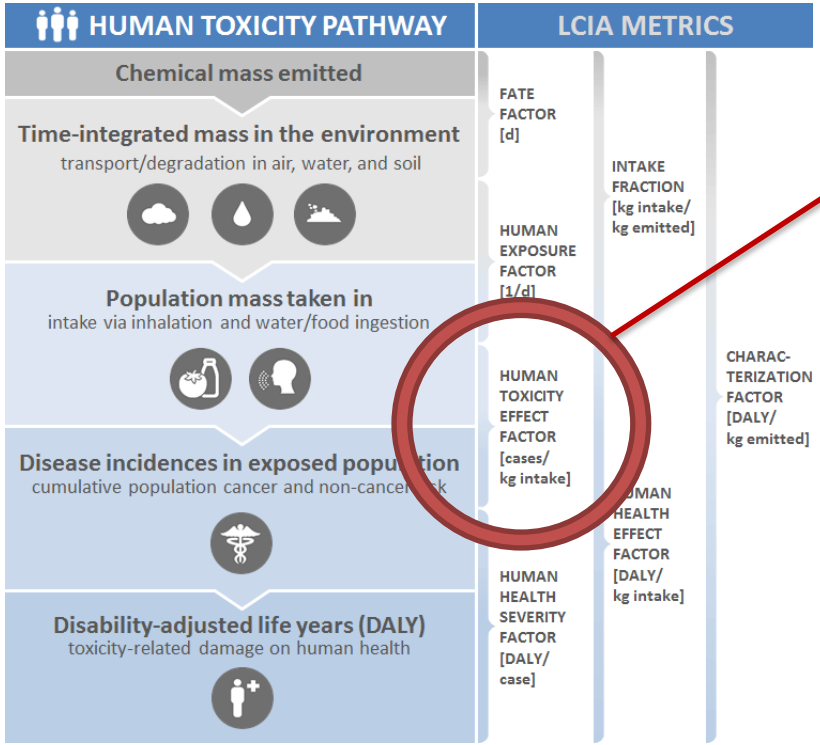
Proceed to "RfD Estimate"



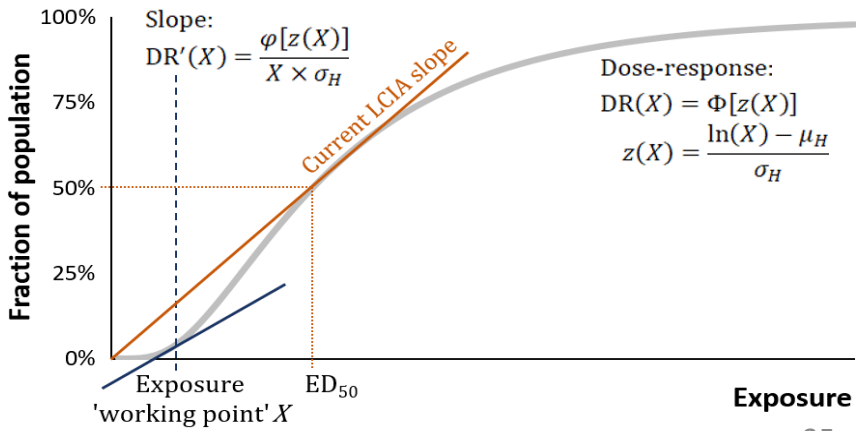


https://benchmarkdose.org/

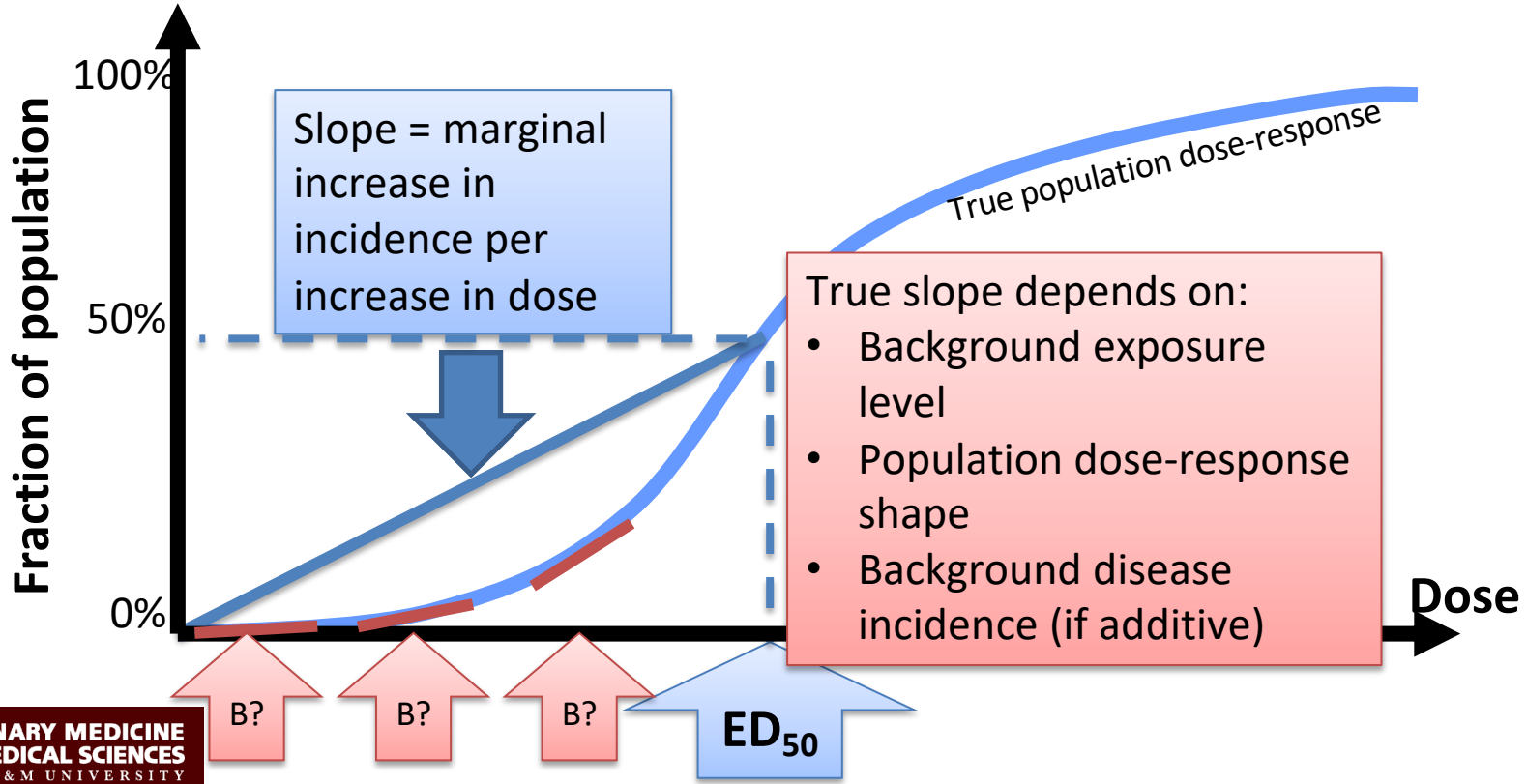
Application to Life-Cycle Impact Assessment



- Primarily used for *comparative risk assessments* of life-cycle impacts
- Requires *slope* of dose-response curve, not a "safe dose" level.



Current LCA Effect Factor (EF) based on linear extrapolation from ED_{50}

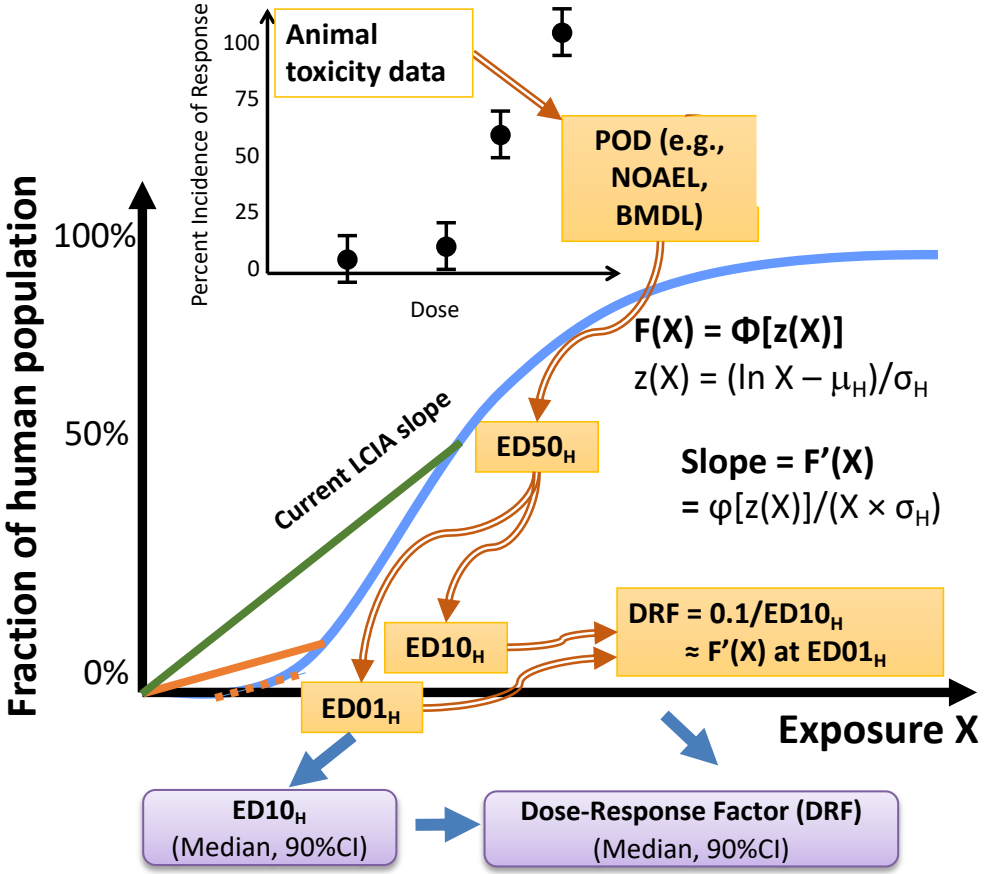


Proposed approach

- LCIA currently uses single point estimate for each factor
- *Linear slope* from the ED10 as a "reasonable" value (with 95% CI) for the *marginal slope* for a wide range of background exposures.
- Consistent with alternate interpretation that marginal slope should be at the background *disease incidence* due to additivity.

Challenges:

- LCIA analyses often include hundreds or thousands of chemicals simultaneously.
- Lack of regulatory toxicity values, or even any *in vivo* data, for vast majority of chemicals.

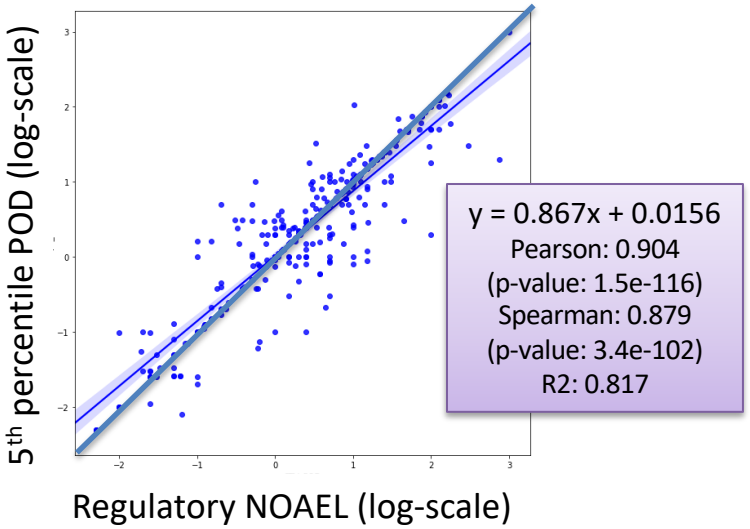


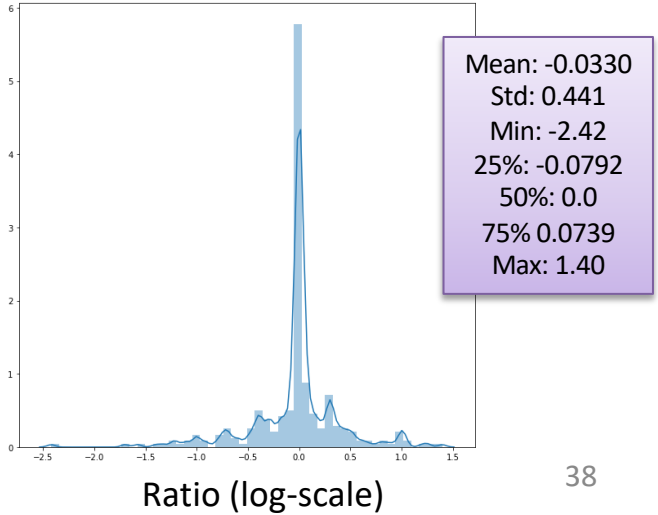
Experimental animal data: Regulatory vs. non-regulatory PODs

- Regulatory toxicity values undergo extensive review process for selecting in vivo PODs.

- In absence of regulatory values, which in vivo POD to use?
  - Minimum?
  - 5th percentile?
  - Median?

Unbiased, but with fat tails





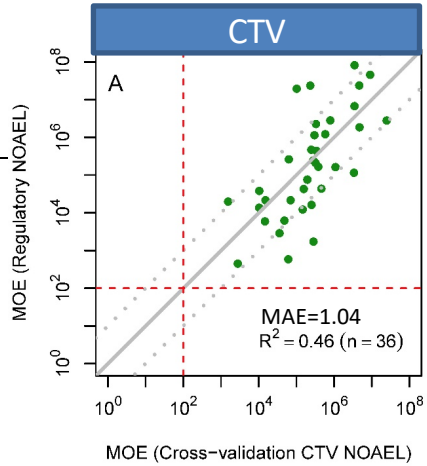
Data Sources: Paul Friedman et al. 2020; Wignall et al. 2018

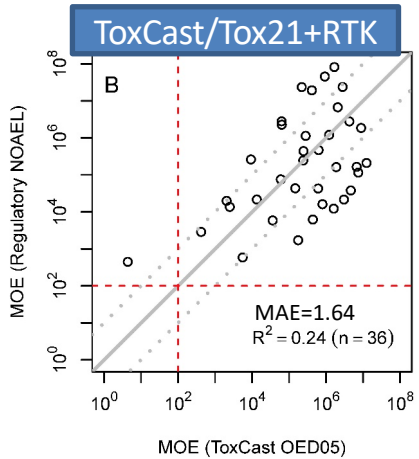
NAMs as PODs: QSAR vs. ToxCast/Tox21

- CTV (Wignall et al. 2018, toxvalue.org) developed a suite of QSAR models for predicting regulatory toxicity values (e.g., NOAELs) in the absence of *in vivo* data.
- ToxCast/Tox21 toxicity values for screening and prioritization based on
  - In vitro high throughput screen (HTS) assays
  - Reverse toxicokinetics (RTK) to convert in vitro concentrations to oral equivalent doses
- Can evaluate performance relative to "gold standard" of published regulatory toxicity values.

CTV (QSAR-based) has
- larger R^2 and
- smaller MAE

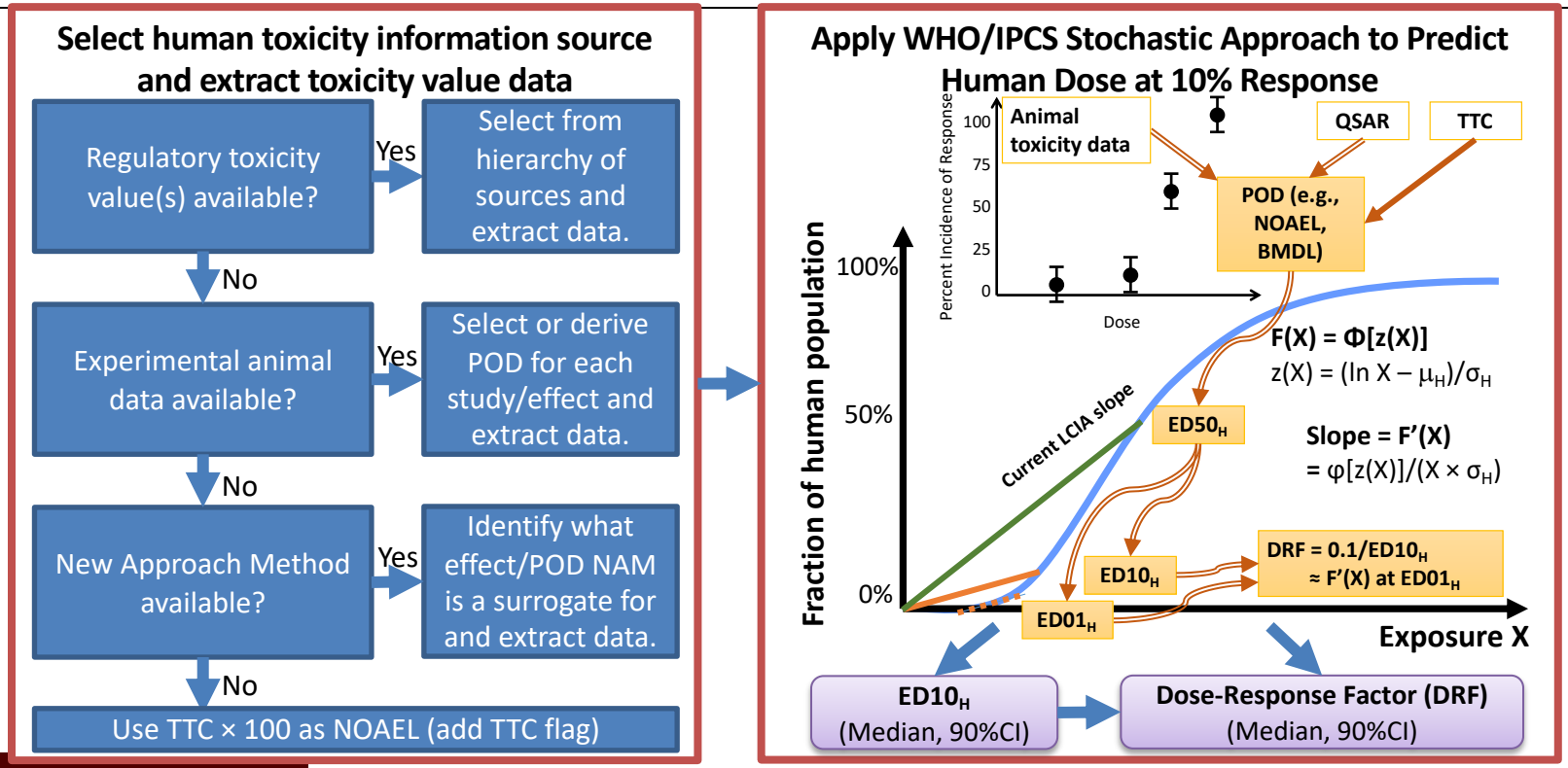
comparted to HTS+RTK-based OEDs



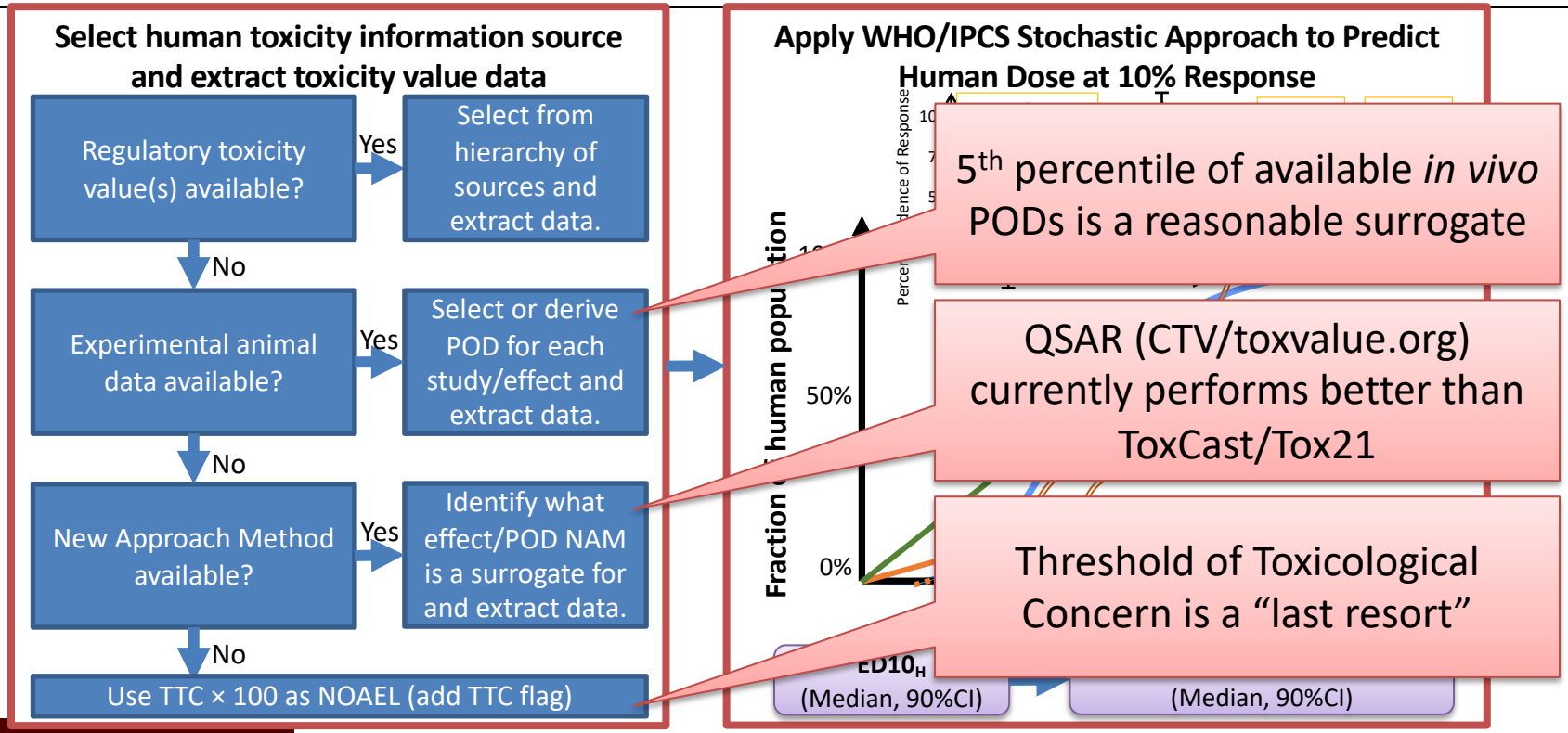


Source: Wignall et al. 2018

Proposed decision-tree flowchart



Proposed decision-tree flowchart



Analysis ongoing as part of UNEP / USETOX project…. Stay tuned!

Outline

- Motivation and background
- Overview of the WHO/IPCS framework
- Recent advances in apply the WHO/IPCS framework
- **Remaining challenges and summary**

Challenges ahead

- Quantification challenges
  - Even animal toxicology data is limited to at most a few thousand chemicals/endpoints – more investigation of use of NAMs needed
  - For benefit-cost analyses, difficulty in assigning economic value to most toxicology endpoints
  - High uncertainty for low levels of incidence

- Institutional and communication challenges
  - Lack of experience/training with probabilistic methods
  - Requires more explicit risk management specification of protection goals (no longer binary)
  - Communicating residual risk & uncertainty is difficult.

Summary

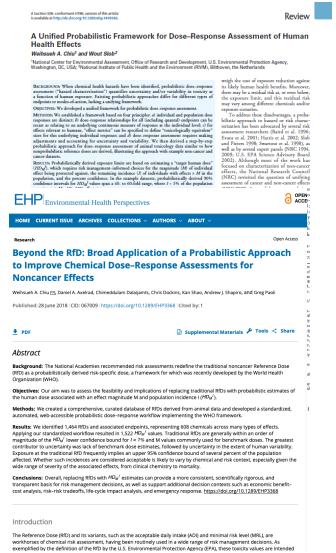
Bottom line: WHO/IPCS probabilistic approaches provides a substantial bridge between toxicology data and a broader suite of risk management decision contexts.

- Replacing RfD with HD_M^I is directly analogous to replacing the NOAEL with the BMDL.
- Approach addresses numerous limitations of current practices.
- Results include point estimates, confidence limits, and continuous dose-response functions.
- Framework can be broadly implemented using existing software and tools.
- Numerous opportunities exist beyond economic benefits analysis for employing more useful dose-response assessments
- Several challenges remain for broader application.

For more information…

Publications

- WHO (2014, 2017 [2nd edition]): Guidance on Evaluating and Expressing Uncertainty in Hazard Assessment. http://www.who.int/ipcs/methods/harmonization/areas/hazard_assessment/en/ (including APROBA spreadsheet)
- Bokkers et al. (2017) APROBA-Plus: A probabilistic tool to evaluate and express uncertainty in hazard characterization and exposure assessment of substances.. FCT, DOI: https://doi.org/10.1016/j.fct.2017.10.038
- Chiu WA & Slob W (2015): A Unified Probabilistic Framework for Dose-Response Assessment of Human Health Effects. EHP, DOI: http://dx.doi.org/10.1289/ehp.1409385
- Chiu WA. (2017) Chemical risk assessment and translation to socio-economic assessments. OECD Environment Working Papers. 2017 March 14; (117). http://dx.doi.org/10.1787/a930054b-en
- Chiu et al. (2018) Beyond the RfD: Broad Application of a Probabilistic Approach to Improve Chemical Dose-Response Assessments for Noncancer Effect. EHP, DOI: https://doi.org/10.1289/EHP3368
- Fantke et al. (2018) Advancements in Life Cycle Human Exposure and Toxicity Characterization. EHP, DOI: https://doi.org/10.1289/EHP3871
- Wignall et al. (2018) Conditional Toxicity Value (CTV) Predictor: An In Silico Approach for Generating Quantitative Risk Estimates for Chemicals. EHP, DOI: https://doi.org/10.1289/EHP2998



Funding:

- TAMU Superfund Research Center (NIH P42 ES027704)
- KS and Associates LLC (via NIH STTR 1R41TR002567-01A1)